

Яламов Г.Ю.

ФОРМИРОВАНИЕ КОНТЕНТА ОБРАЗОВАТЕЛЬНЫХ ВЕБ-САЙТОВ

*ФГБНУ «Институт управления образованием РАО», г. Москва,
geo@portalsga.ru*

Yalamov G.Yu.

FORMATION OF CONTENT FOR EDUCATIONAL WEBSITES

The Federal State Budgetary Scientific Institution «Institute of Management of Education of The Russian Academy of Education», Moscow, geo@portalsga.ru

Рассмотрены способы формирования контента и инфраструктуры образовательных сайтов и сайтов научных организаций с учетом анализа подробных пассивных статистических данных обращения пользователей к его информационным ресурсам. Приведен пример анализа статистики Российского портала информатизации образования.

Ключевые слова: образовательный веб-сайт, веб-сайт научной организации, портал, контент, статистика посещений, анализ статистики.

Methods of formation of content and infrastructure of websites educational and websites research of a scientific organizations are considered, taking into account the analysis of detailed passive statistics of users' access to their information resources. An example of the analysis of statistics of the Russian portal of informatization of education is given.

Keywords: educational website, website research of a scientific organization, portal, content, attendance statistics, analysis of statistics.

В настоящее время поддержка, развитие и продвижение собственного веб-сайта является важной составляющей деятельности образовательных и научных организаций. Во многом это обусловлено тем, что количество упоминаний в СМИ об организации и данные о посещаемости их официальных веб-ресурсов являются одними из критериев оценки деятельности образовательных и научных организаций Министерства образования и науки РФ [1]. Кроме того, российские университеты

и научные учреждения (центры) входят как в международные вебметрические рейтинги (такие, как Consejo Superior de Investigaciones Científicas, Cybermetrics Lab и др.), так и в отечественные (рейтинг ИПМИ КарНЦ РАН). Заметим, что последний учитывает в том числе и данные поисковой системы Яндекс, занимающий, как известно, первое место в России среди прочих поисковых систем.

Таким образом, для успешного поискового продвижения веб-сайта возникает необходимость оценить его популярность, активность и предпочтения его пользователей в отношении предлагаемых на сайте информационных материалов и документов, страниц сайта в целом. Для этого будет полезен анализ данных, предоставляемых онлайн сервисами (системами сбора статистической информации) Hotlog, Liveinternet, Яндекс.Метрика, Google Analytics и др. Практическая ценность такого анализа состоит в том, что он дает основания для реализации практических шагов по поддержанию эффективного функционирования веб-ресурса, позволяет сформулировать практические выводы и рекомендации по дальнейшему развитию и актуализации его информационных ресурсов, выявляет недостатки.

Но методы сбора статистики в данных системах представляются нам неполными, требующими дополнительных исследований. Это особенно заметно в отношении веб-сайтов, направленных на сбор и накопление информации, имеющих достаточно большой массив документов образовательного и научного характера, предоставляемых пользователям для свободного скачивания. Это можно объяснить следующими факторами:

1. Вышеуказанные системы ведут учет активности пользователей только на страницах html-формата, на которых установлены коды счетчиков. Но многие образовательные и научные веб-сайты (далее сайты) созданы на других платформах, страницы которых или вообще не имеют html-кода, или имеют лишь его фрагменты. Таким образом, не все страницы этих сайтов охвачены статистикой (во многих случаях их большая часть).
2. Многие сайты имеют динамические веб-страницы, которые генерируются в момент запроса браузером. Статистика по таким веб-страницам также не ведется.

3. Онлайн сервисы не предоставляют статистической информации по запросам и просмотрам пользователями конкретных документов (форматов doc, docx, pdf, ps и др.) составляющими основной контент многих сайтов. Тем не менее, такая информация очень полезна при формировании контента сайтов, актуализации их материалов.
4. Не ведется учет пользователей, обратившиеся к документам сайтов по запросу с поисковых систем по прямой ссылке, минуя страницы сайтов.

Поэтому, для получения развернутого и полного представления об активности пользователей в отношении сайтов, их количестве, популярных страницах, документах и материалах, запрошенных пользователями, степени соответствия тематик сайтов этим запросам, доминирующих интересах пользователей, рекомендуется проводить структурный анализ статистических данных сайтов с использованием статистических данных, записанных в лог-файлы веб-серверов сайтов. Такой метод сбора статистических данных некоторые авторы называют пассивным, так как он осуществляется на стороне веб-сервера, силами сотрудников организации владельца сайта [2].

Несмотря на некоторые погрешности, лог-файлы веб-сервера являются самым полным источником статистики посещаемости и запросов к ресурсам сайтов. Основное внимание при обработке записей лог-файла веб-сервера сайтов мы рекомендуем акцентировать внимание на выявлении статистических данных о среднесуточном количестве уникальных посетителей сайтов, запросах, загрузках и просмотрах документов и материалов базы данных сайтов, как при непосредственных посещениях страниц сайта, так и при прямых запросах поисковых систем к документам сайтов. Это позволит идти в правильном направлении при формировании, редактировании и актуализации контента сайтов.

При выполнении анализа можно использовать разные инструментальные средства, в том числе и программы лог-анализаторы. Одной из таких программ является система статистики Web Log expert Std/Pro – современный лог анализатор, который позволяет получить интересующую нас статистическую информацию в контексте данной статьи [3]. Приведем пример исследования некоторых статистических данных Российского портала информатизации образования (далее РПИО), главная страница которого расположена по адресу: <http://portalsga.ru>. При

анализе этих данных был использован отчет, сгенерированный одной из последних версий программы Web Log expert Pro v.9.3 (далее программа). Нас в первую очередь интересовала статистика по запросам и загрузкам пользователями полнотекстовых документов и материалов портала, так как наиболее важной частью сайтов, определяющей цель их создания и поддержки является их информационный ресурс.

Для получения объективных статистических показателей программой были игнорированы IP-адреса участников проекта (такая возможность предусмотрена функционалом программы).

Общая статистика посещений РПИО за указанный период представлена в таблице 1.

Таблица 1. Общие показатели статистики РПИО (16.04.17-21.04.17)

Посетители	
Среднесуточное число посетителей	348
Среднесуточное число уникальных посетителей	311
Общее число посещений	2088
Общее число уникальных посетителей	1869
Запросы к ресурсам РПИО	
Суммарное число запросов	19245
Среднесуточное число запросов	3207
Число посещений поисковыми роботами	3153
Среднесуточное число запросов на 1 посетителя	771
Отказ запросов	1070
Среднесуточное число скачиваемых полнотекстовых документов и материалов РПИО	285
Просмотры страниц РПИО	
Общее кол-во просмотров	1477
Среднесуточное кол-во просмотров	246
Среднесуточное кол-во просмотров на 1 посетителя	1,41

Внесем некоторые пояснения.

Понятие посещения (также сеанс, сессия, просмотр) расценивается программой как период взаимодействия между браузером посетителя и определенным сайтом, завершающийся

при закрытии окна браузера или неактивности пользователя на этом сайте за определенный период времени. Например, в контексте отчетов Google Analytics сеанс считается завершенным, если пользователь не проявляет активности на сайте в течение 30 минут. Мы придерживаемся данному определению.

Программа включает в число посетителей и тех пользователей, которые загрузили документы портала по прямому запросу с поисковых систем (Яндекс, Google и др).

Уникальные посетители – это учитываемые только один раз в сутки посетители сайта за указанный период времени. Определяется программой по IP-адресу.

Необходимо заметить, что количество уникальных посетителей за указанный период времени, полученное с помощью сервиса онлайн-статистики Hotlog равно 167, что составляет всего 9% от того же показателя, полученного при нашем анализе. Существенно разнятся и другие показатели.

Далее приведем данные об активности пользователей относительно полнотекстовых документов учебного и научно-методического характера, представленных на портале для свободного просмотра и скачивания (таблица 2).

Таблица 2. Пользовательские загрузки наиболее популярных документов РПИО (16.04.17-21.04.17)

№	Ссылка на документ	Число просмотров	Число загрузок	Число отказов загрузки	Объем данных (КВ)
1	http://www.portalsga.ru/data/3112.docx	405	370	0	1025469
2	http://www.portalsga.ru/data/2394.docx	156	139	0	18359
3	http://www.portalsga.ru/data/2309.docx	124	107	0	18180
4	http://www.portalsga.ru/data/2902.doc	100	90	2	166284
5	http://www.portalsga.ru/data/2475.docx	123	79	29	88070
6	http://www.portalsga.ru/data/2311.docx	83	66	0	4855

7	http://www.portalsga.ru/data/2272.docx	64	58	0	6368
8	http://www.portalsga.ru/data/2386.docx	66	55	0	13636
9	http://www.portalsga.ru/data/2297.docx	48	46	0	25145
10	http://www.portalsga.ru/data/2294.docx	45	43	0	18070

Здесь мы ограничились представлением данных о первых 10-ти документах, полученных из отчета программы за неделю, но программа позволяет получить отчет по всем документам практически за любой период времени. Мы рекомендуем проанализировать полный отчет за период времени не менее месяца, акцентировав внимание на следующем:

- Тематика, характер (учебник, учебное пособие, учебно-методическое пособие, научная статья, автореферат и др.) первых 10-25 документов в отчете программы.
- Соотношение просмотров и загрузок по каждому из этих документов.
- Количество отказов загрузки (неполная загрузка) по каждому из этих документов.
- Формат каждого из этих документов.
- Отсутствие статистики или сравнительно низкая статистика по определенным документам (за длительный период, 30-60 суток).

Результаты такого анализа дадут правильный ориентир для дальнейшего формирования, редактирования и актуализации контента сайта в части полнотекстовых документов, повышения обеспеченности сайтов востребованными документами, удобства их использования.

Представляют интерес и данные по статистике обращений пользователей к страницам портала. На рисунке 1 представлены такие данные по 10-ти наиболее популярным его страницам (здесь Visitors – посетители). В нашем случае эти данные носят относительный характер, так как основной контент портала это полнотекстовые документы научно-педагогического, учебно-

методического, научно-популярного, учебного, справочного, нормативно-инструктивного и организационного характера (всего более 1000 документов). Тем не менее, такие данные должны учитываться при проведении работ по развитию и продвижению сайтов.

Безусловно, мы не рекомендуем ограничиваться анализом лишь тех данных, которые приведены в статье. В общем случае необходим учет всех показателей статистики. Кроме того, с целью

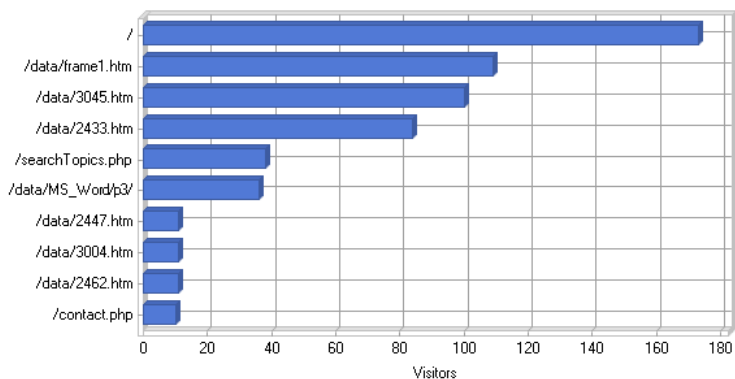


Рис 1. 10 наиболее популярных страниц РПИО

поддержки, успешного развития и продвижения сайтов при формировании и управлении их контентом и инфраструктурой рекомендуется учитывать и следующие факторы [4]:

- способы и формы представления информации на сайтах;
- результаты мониторинга и анализа содержания других сетевых информационных ресурсов, близких по тематике к контенту сайтов;
- отзывы на качество информационных материалов от пользователей сайтов, образовательных, научных организаций и обществ, органов управления образованием в электронных Интернет-публикациях и СМИ;

Литература

1. Кабакова Е. А. Веб-сайт научно-исследовательского учреждения: наполнение, посетители, развитие [Электронный ресурс] / Е. А. Кабакова, В. С. Усков // Вопросы

- территориального развития. 2014. – № 3. – Режим доступа: <http://vtr.vscs.ac.ru/article/1396/full>.
2. Skorodumov P. V., Kholodev A. Yu. Analiz podkhodov i instrumental'nykh sredstv analiza statistiki poseshcheniya veb-saita nauchnoi organizatsii [Analysis of approaches and tools to analyze the statistics of visits to the web site of scientific organizations]. Voprosy territorial'nogo razvitiya [Issues of territorial development], 2015, no. 9. Available at: <http://vtr.vscs.ac.ru/article/1650/full>.
 3. Яламов Г. Ю. Об использовании ресурсов Всероссийского студенческого информационного портала/ Г. Ю. Яламов // ИНФОСЕЛЬШ-2009: Труды VI Всероссийского научно-методического симпозиума. – 2009. – С. 544-552.
 4. Ваграменко Я. А. Архитектура и формирование контента информационной системы для обеспечения научно-образовательной и воспитательной деятельности в молодежной среде/ Я. А. Ваграменко, Г. Ю. Яламов // Электронное периодическое издание Информационная среда образования и науки. – 2014. – №19. – С. 5-47.